Lab 4 Guide

Retrieval Practice and Learning

This week's lab revisits the **pivot tables** introduced in Lab 2B. Evidence of Discrimination. Recall that pandas pivot tables take 3 main parameters:

- values - the dependent variable(s) of interest
- index - the independent variable(s) of interest
- aggfunc - how you want to summarize your data (e.g. sum, mean, count, min/max)

In lab 3B, we saw that pivot tables can also take another parameter called columns, which **groups** data based on the different values of the parameter.

Here are some questions to consider for **Question 1**:
- What would the x-axis (values) and y-axis (index) be?
- Which aggregation function best summarizes the overall scores from both groups?

Here are some questions to consider for **Question 2**:
- What are the TS.1, TS.2, and TS.avg columns?
- Which one would be the most appropriate to use for the y-axis?

For **Question 3**, you can **plot** values from Questions 1 and 2 **side by side** by passing in both columns as the values parameter:

$$... , values = [\ 'col1' \ , \ 'col2' \ ] \ , ...$$

You may find crosstab() more useful than pivot_table() for **Question 4**.